Arh hig rada toksikol 1998;49:307–317

ORIGINAL SCIENTIFIC PAPER

DEVELOPMENT IN HARMONISATION OF PROFICIENCY TESTING (FOR VAPOURS, GASSES, AND DUSTS) IN THE EUROPEAN UNION

PETER STACEY AND BARRY TYLEE

Health and Safety Laboratory, Sheffield, United Kingdom

Received 7 December 1998

A network has been established of the 11 major proficiency schemes in the European Union concerned with the occupational hygiene and environmental analyses of chemicals and dusts in the air. A comparison of all the schemes was carried out and a compendium is being produced. This will allow users of the schemes such as testing laboratories, customers, and regulatory bodies to choose the scheme that is most suited to their purpose. All schemes have been compared with the revised ISO Guide 43, published in 1997. The performance statistics in most schemes conform to the criteria in European Standard EN 482 that define the acceptability limits for overall uncertainty in measurement. However, the performance statistics and assessment strategies of the different schemes vary. While many of the schemes supply similar sample material such as lead on filters and benzene on charcoal, there are a number of sample types that many schemes would like to introduce. However, it would be uneconomic to do this on a national basis and the network is developing procedures to introduce them throughout the member countries. Additionally, there are countries that have no schemes at present and may wish to introduce them. The network will provide a framework to help set up schemes in these areas.

Key words:
air analysis, air monitoring, chemicals, dusts, environmental analysis, European network, quality assurance, workplace analysis

In the workplace, chemical agents present as gases, vapours, or particles such as dust, fumes, fibres, or mists, likely to be hazardous to health are sampled and collected by drawing air through various collection media. For hazardous metals in the air, such as lead or cadmium, the medium is a membrane or glass fibre filter. For organic

Presented at the AOAC INTERNATIONAL Central Europe Subsection 5th International Symposium on Interpretation of Chemical, Microbiological and Biological Results and the Role of Proficiency Testing in Accreditation of Laboratories, Varaždin, Croatia, 21–23 October 1998.

compounds, such as benzene or xylene, the medium may be charcoal granules contained in a glass tube or synthetic sorbents in stainless steel tubes. The organic solvents may either be chemically desorbed using a suitable solvent or thermally desorbed by heating the tube. The metals on the filters are dissolved into solutions and analysed later by an appropriate technique. If uncontrolled, the errors in these analyses can be very large. Such measurements need to be accurate, or may otherwise affect the health of the employees or commit a company to costly improvements that are not required. Proficiency-testing schemes have arisen to meet the need to assess the quality of measurements made on airborne contaminants found in the workplace.

Schemes exist in several member states of the European Community such as Belgium, the Netherlands, Spain, France, Denmark, and the United Kingdom. Several member states require specific levels of performance to be obtained in their own national proficiency testing schemes for analytes such as lead, benzene, or asbestos fibres. The consequence of an organisation not achieving this performance is that they are not allowed to perform these analyses.

It is the common experience of the proficiency testing schemes in Europe that participation in them is very effective in improving the analytical accuracy of laboratories. As an example, Figure 1 shows the experience of the Workplace Analysis Scheme for Proficiency (WASP) which is the United Kingdom scheme for occupational and environmental air analysis.

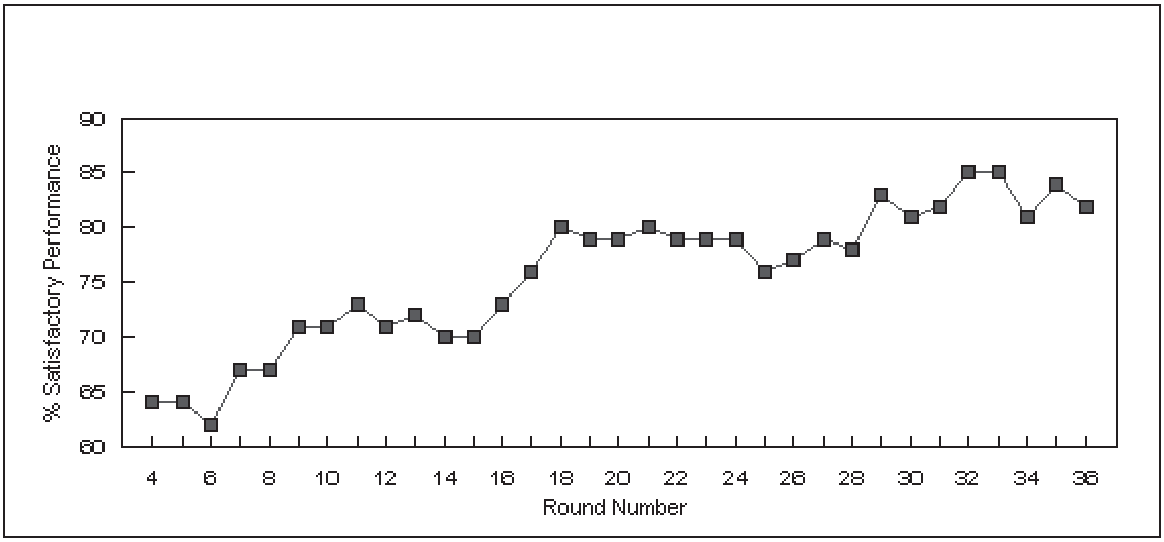

Figure 1 *Improvement in per cent satisfactory performance in the Workplace Analysis Scheme for Proficiency*

When WASP started, almost one in three laboratories reported unsatisfactory results. This implies that 1 in 3 laboratories would have reported results to their customers that would be considered inaccurate. By round 40, the situation is much better with more than 80% of participants achieving a satisfactory performance, which is generally set at 8% coefficient of variation. However, the WASP samples are pure analytes without interferences associated with the »true« working environment and a laboratory's performance is expected to be worse when analysing real samples.

The European Community is establishing directives that, among other things, set specific limits for occupational and environmental exposure. An example of this is the recent adoption by the European Community of the European Air Quality Framework and first daughter directives covering the emission of SO_2,NO_2, PM_{10}, and lead in the environment. It is important that measurements taken in support of these European directives are accurate and of a comparable standard so that organisations in different member states are able to operate across national boundaries and the results obtained are fit for purpose, and easily understood by different regulatory and accreditation organisations.

Schemes exist in other European countries outside the European Community such as Poland. Other national laboratories that do not operate proficiency testing schemes, such as Ireland, Switzerland, and Portugal, participate in other country's schemes to monitor their own analytical performance.

BACKGROUND

A network of proficiency testing schemes from member states of the European Community involved in the analysis of chemical agents in the environment and workplace was established in November 1996 with the support of the European Commission. The network includes 11 organisations from 9 member states of the European Union (Table 1). The aims of this network are:

1. To act as a steering group and umbrella organisation for the development and harmonisation of proficiency testing schemes in the European Union.

2. To develop research initiatives, if required by the network, to improve the quality of proficiency testing standards.

3. To develop a joint approach to improve the quality of analytical performance in occupational hygiene and environmental measurements in industry.

The network is co-ordinated by the Health and Safety Laboratory in the United Kingdom and receives financial support from the Commission. Until the first meeting of the network in 1997, proficiency testing schemes in this field in the European Union had developed as separate entities. Each of the schemes has developed in their own way to suit different national emphases and concerns. These include matters concerning their relationship with legislation, national status, types and quality of proficiency testing material, analytical range, and priorities concerning future development.

APPROACH OF NATIONAL SCHEMES

The relationship between proficiency testing schemes and national legislation is different in each country. In the United Kingdom the WASP scheme was established in support of national legislation known as the Control of Substance Hazardous to Health (COSHH) 1998. This legislation obliges employers to regularly monitor the environment in which their employees work if they are exposed to hazardous substances. The scheme is

Table 1 *Members of the European network*

Institute	Contact	Country
The Health and Safety Laboratory (HSL) Broad Lane Sheffield S3 7HQ	Mr. Barry Tylee	United Kingdom
Institute of Occupational Medicine (IOM) 8, Roxburgh Place Edinburgh EH8 9SU	Dr. Nigel Crawford	United Kingdom
National Institute of Occupational Health (STAMI) Gydas vei 8 P. O. Box 8149 DEP N-0033 OSLO	Dr. Yngvar Thomassen	Norway
Institute National De Rechèrche Et De Securitè (INRS) 27 Avenue de Bourgogne 54501 VANDOEUVRE	Dr. Sandino	France
Laboratorium voor Industrielle Toxicologie (LIT) Belliardstraat 51 Brussels B-1040	Dr. Rodger Grosjean	Belgium
Oulu Regional Institute of Occupational Health (ORIOH) Aapistie 1 FIN - 90220 Oulu	Prof. Lauri Pyy	Finland
Instituto Nacional De Seguridad E Higiene En El Trabajo Centro Nacional De Verificacion De Maquinaria Camino De La Dinamita S/N 48903 Barakaldo (VIZCAYA)	Dr. M.J. Quintana	Spain
Vlaamse Instelling voor Technologisch Onderzoek (VITO) Boeretang 200 Mol 2400	Dr. Ing. Goelen	Belgium
Nederlands Mettinstituut (NMi) Van Swinden Laboratorium B. V. P. O. Box 6011 Schoemakerstraat 97 Delft 2600 Ar Zuid- HOLLAND	Dr. Th. L. Hafkenscheid	Netherlands
Berufsgenossenschaftliches Institut für Arbeitssicherheit (BIA) Alte Heerstrasse 111 Sankt Augustin 53757 Nordrhein-Westfalen (NRW)	Dr. Breuer Dietmar	Germany
Arbejdstilsynet (AMI) Lersoe Parkalle 105 Copenhagen DK-2100	Dr. Christensen	Denmark

Acknowledgement: This network receives financial support from the European Commission's Standards Measurement and Testing program.

voluntary and relies on the customers of the laboratories requesting evidence of quality assurance. Most of the schemes in the network are voluntary, but in some countries such as France or Spain, laboratories must achieve a specific level of performance for the analysis of analytes such as lead, benzene, or asbestos fibre counting in the national scheme. Otherwise they are prevented from operating in that country. Other schemes have strong links with their national system of laboratory accreditation.

The main factors contributing to the overall uncertainty of a measurement are the sampling and the chemical analysis. These areas are outlined in Figure 2.

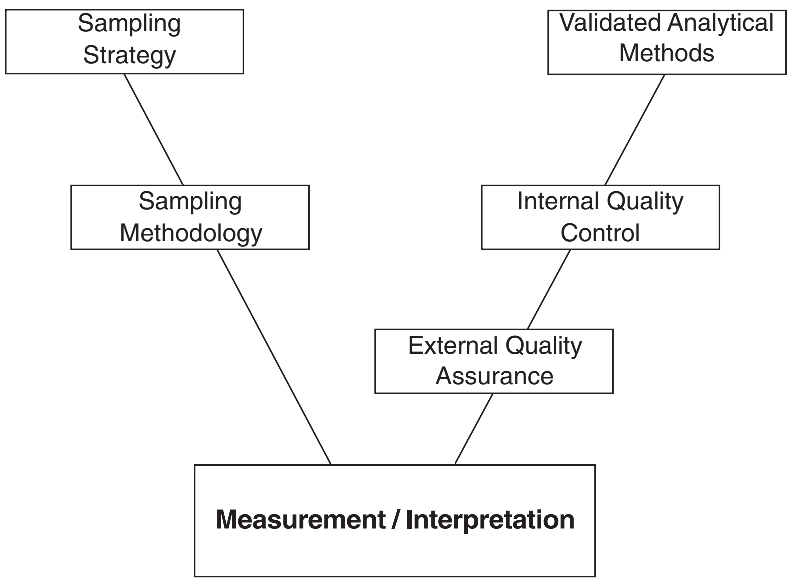

Figure 2 *Factors that contribute to the quality of an analytical measurement*

Many national regulatory authorities try to ensure that accurate measurements are taken in the workplace by investing in the development of validated methods, and providing guidance concerning the sampling strategy and methodology.

In the workplace monitoring field, the European Standard EN 482 (1) states that contribution of the overall errors for both sampling and analysis should not be greater than ±30% around the exposure limit value and ±50% at half the limit value. In Belgium, the Flemish regional government has incorporated this standard into the Vlarem legislation and a proficiency testing scheme in this country attempts to assess the overall uncertainty by inviting its participants to sample an artificially generated vapour using their own equipment. The federal ministry together with the Vlaamse Instelling voor Technologisch Onderzoek (VITO) institute organises these sampling exercises for a large number of organic substances and participation is compulsory for laboratories that wish to obtain formal recognition or accreditation. A similar policy exists in France for the authorisation of laboratories for the analysis of benzene.

This method has a disadvantage as it is expensive for participants to spend a day at a central location and these exercises are not held as frequently as other schemes. However, it is one of the few schemes where participants are able to evaluate the proficiency of their sampling. Sampling is often the most poorly controlled factor contributing to the overall uncertainty of a result. Most schemes concentrate on attempting to control the analytical proficiency by regularly sending participants standard samples. In the UK WASP scheme, the EN 482 criteria has been interpreted in the following way; the contribution of the errors in analysis is expressed mathematically as:

$$\sigma_{TOTAL} = \sqrt{\sigma_a^2 + \sigma_s^2}$$

where σ_a^2 is the analytical contribution and σ_s^2 is the sampling contribution. If it is assumed that the sampling contribution is large and constant, there will be a point when any further control of the analytical contribution will not be effective. Mathematically this is about 1/4 of the overall uncertainty. If an overall variation of ±30% is considered, then the σ_a^2 level is about 8%. Most performance limits in WASP are set at this level if it is analytically achievable.

PROGRESS TOWARDS HARMONISATION

The network undertook an exercise to compare each scheme to the international standard Proficiency Testing by Interlaboratory Comparisons (ISO) 43 (2). Generally, most schemes, although they may vary considerably in their approach, compare favourably with the recommendations made in ISO 43. Each scheme has agreed to accept participants from other countries. However, many differences between the schemes remain unresolved.

Differences in protocols

The most common way of providing external quality assurance to laboratories is to provide proficiency testing samples on a regular basis. However, the protocols of each scheme differs in a number of ways:

1. not all schemes send the same analytes to participants in every round.
2. some schemes involve a qualitative aspect as well as the quantitative aspect.
3. the number of samples sent to participants each year may vary from 3 to 32.
4. the number of rounds may vary from once every 3 years to 4 times a year.
5. not every scheme produces a performance statistic that assesses the performance of its participants over more than one round.
6. the level set by the schemes to determine competence can vary from 6 or 8%, or two standard deviations of the population (excluding outliers). Fibre counting schemes can have much wider limits.

Statistical protocols

The two most common statistics used in the proficiency testing schemes are the bias or q score:

$$\left(\frac{x-\mu}{\mu}\right)$$

where x is the participant's result and μ is the assigned value; and the z score:

$$\left(\frac{x-\mu}{s}\right)$$

where s is a measure of variance determined by expert consensus.

These two scores are very similar; however, not all the schemes use these statistics and the criteria to assess whether satisfactory performance is obtained can vary 6–8% from a standard distribution of data.

An exercise to compare the statistical protocols for those schemes that involve benzene highlighted the differences between those that regularly send the same analytes to their participants and those that send different analytes in each round. Generally, those schemes that regularly send participants the same analytes produce a performance score that takes account of a laboratory's performance over a number of rounds, whereas the schemes that send participants different analytes in each round generally produce a statistic that takes account of the performance of a laboratory in a single round. Most laboratories have erratic performance over a period of time. Figure 3 shows a typical example of a laboratory analysing benzene in the air in the WASP scheme.

Initially, the laboratory's performance is very erratic with some very high results. As it continues in the scheme it manages to control these errors. A scheme that uses a cumulative score would rate the laboratory differently to a scheme that produces a single statistic in each round. The best ranked correlation of performance scores between schemes was obtained from those that use a performance statistic that takes

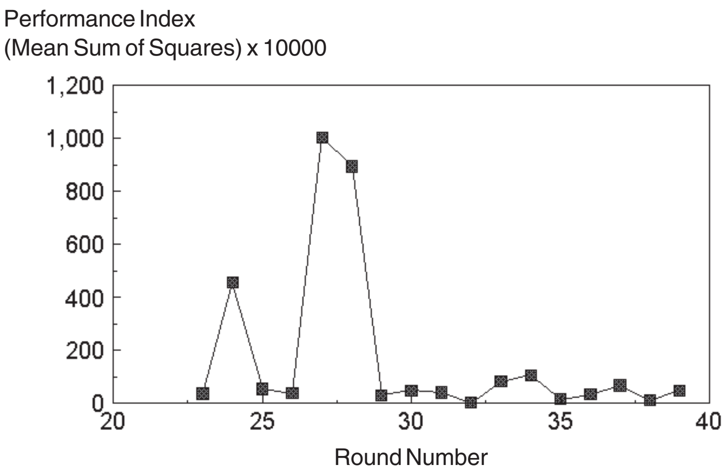

Figure 3 *A typical example of a laboratory with erratic bias – analysis of benzene on Tenax*

into account a laboratory's performance over a number of rounds. The other schemes, generally, were only able to classify a laboratory as having unsatisfactory performance if it had obtained a poor result in a single round, perhaps by chance, or had a consistently poor bias.

In order to introduce some commonality, the network members have agreed to introduce the mean sum of differences squared statistic to operate alongside their existing performance indicator.

$$\frac{\sum_1^n (x-\mu)^2}{n}$$

This statistic is the deviance of the participant's results from the assigned value and is similar to the variance statistic if the data is normally distributed and the assigned value is the mean.

The schemes have also agreed to determine the uncertainty of the assigned values in order to give participants a better understanding of their performance rating.

Development of »realistic« standards

It is important that the proficiency testing material should be as representative as possible of the »true« environment in which the analyte of interest is usually found. That is, they should present to analysts the same type of analytical problems that would be found in real samples. Until recently, the production of »realistic« type samples was prevented because the development costs were considered too large. The only exceptions to this were the national schemes to measure fibres in the air. Analytes such as benzene and toluene were mixed with solvent and injected directly onto the sorbent contained in the tubes, or solutions of metal salts of lead and cadmium were spotted onto filters. Some schemes have been able to take advantage of technology developed as part of the European Communities, Standards, Measurement, and Testing program to produce certified reference materials, and use this to produce proficiency testing standards to the same quality. However, these standards are often very pure analytes on clean substrates and are free from any interferences. This is obviously not the situation when real samples are taken.

In the case of the organic compounds (3), the technology generates a vapour in a chamber. The amount on the sorbent can be calculated theoretically since the volume of air through the system and the weight loss of analytes are known and are traceable to secondary national standards. The vapour generation and accurate sampling removes the need to use solvent solution, often diluted with methanol, in the standard manufacturing process and many different types of mixtures of vapours can be generated and absorbed onto the medium in accurately known amounts. This adds to the complexity of the samples that can be produced.

The samples used in the WASP scheme are usually pure analytes on clean substrates. Recently, participants were given the option of analysing an extra set of samples of benzene, toluene, and *m*-xylene on charcoal with two added components, iso-butyl acetate and ethyl benzene. These components were specifically chosen because they would elute near the toluene and *m*-xylene peaks, if participants used a non-polar column. The histograms in Figure 4 demonstrate the effect of increasing the complexity of standards sent to participants. The first histogram is the distribution obtained by participants analysing the regular samples of pure analyte and a clean substrate that are used in the WASP scheme. The second histogram is the distribution of participants' results for the same analyte with the added components. This distribution is wider and has a negative bias.

For the metal analytes, a method of producing replicate standards from aerosols of fume, dusts and powders has been developed by the National Institute of Occupational Health (SAMTI) in Norway (4). This is known as a multi-port sampler (5). It consists of a large chamber into which the aerosol is drawn. At the top of the cylindrical chamber are a number of filter heads; behind each head is a critical orifice. When a vacuum is applied behind the critical orifice its size regulates the rate at which the air is drawn through the filter. The size of the critical orifices is machined and polished in a consistent way and replicate samples from aerosols of dust or fume can be produced with reasonable precision.

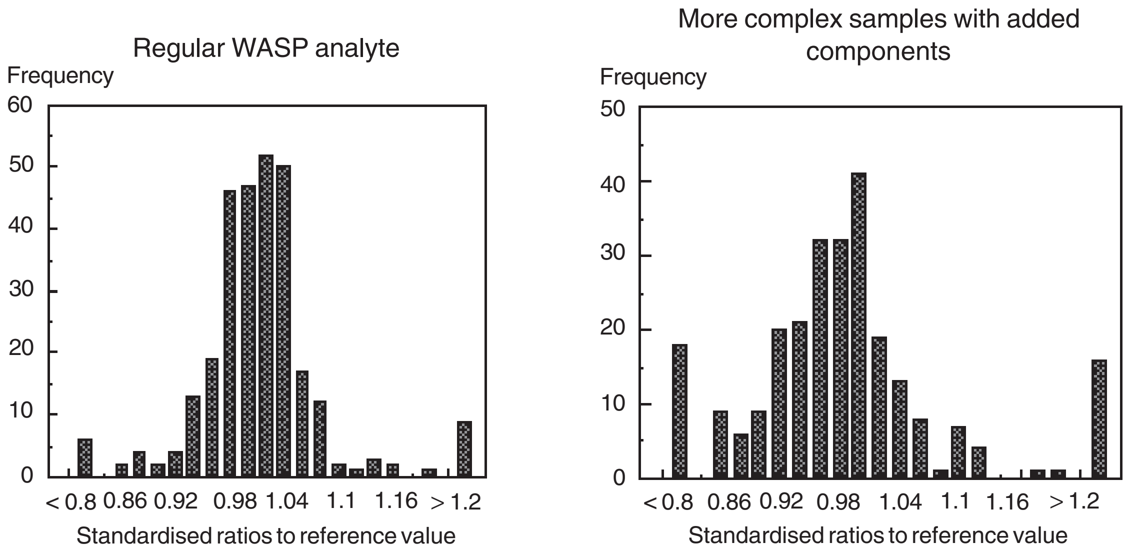

Figure 4 *Comparison of analysis of toluene on charcoal*

NEW ANALYTES

The network is beginning to develop a co-ordinated strategy and identify analytes or sample types that could be introduced Europewide with each scheme acting as a national centre. The network is considering possible introduction of hexavalent chromium from welding fume as a new European initiative. Analysis of metals in welding fume is of concern since it is considered difficult using current methods available in industry, and studies have shown that laboratories have difficulty in obtaining accurate results. The table below shows the distribution of results obtained during a trial introduction of chromium in welding fume samples into the WASP scheme. Figure 5 indicates that participants who analysed this sample generally have a negative bias. It is possible that the situation is slightly better since not all WASP participants who analysed the trial sample would analyse these samples routinely or use a validated method. However, selected national laboratories involved in the WASP scheme also had similar problems when analysing this sample.

The introduction of this analyte on a Europewide basis is useful since the numbers of laboratories participating at a national level may not be large enough to support the introduction of this analyte. Other analytes considered are NO_2 on diffusive tubes, oil mist, and quartz on filters.

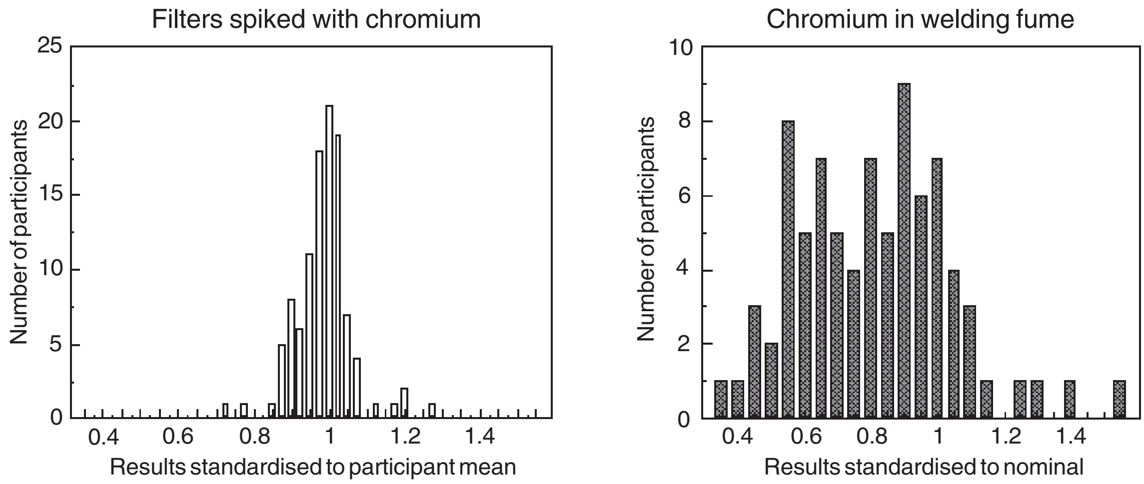

Figure 5 *WASP proficiency testing results, illustrating the difficulty of analysing real type samples*

CONCLUSION

Co-operation between the schemes has led to a greater understanding of each national approach and a process of harmonisation has started. The opening of the schemes will provide a greater choice for participants in Europe who will be able to identify a scheme appropriate to their needs. The network has also facilitated a sharing of technology and an interest of the network members to developing new sample types Europewide.

REFERENCES

1. *European Committee for Standardization (CEN)*. Workplace atmospheres – General requirements for the performance of procedures for the measurement of chemical agents. EN 482. Brussels: CEN, 1994.
2. *International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC)*. Guide 43: Proficiency testing by interlaboratory comparisons. 2nd edition. Geneva: ISO, 1997.
3. *Goelen E, Lambrechts M, Geyskens F, Rymen T*. Development and performance characteristics of a capillary dosage unit with an *in situ* weight sensor for the preparation of known amounts of gaseous VOC's in air. Int J Environ Anal Chem 1992;47:217–25
4. *Bjørgo R, Fjeldstad P, Nagy K, Thommassen Y*. Parallelfiltre fra arbeidsatmosfæren i aluminiumsindustrien for interlaboratoriekontroll av analyse på PAH og flurid. Oslo: Institute of Occupational Health, 1980.
5. *Bonnevie Anglov JT, Holst E, Dyg S, Christensen JM*. Iron, manganese, copper, and titanium in welding fume dust on filters for internal and external quality assurance. Fresenius J Anal Chem 1993;345:335–9.

Sažetak

USKLAĐIVANJE TESTOVA VRSNOSTI ZA AKREDITIRANJE LABORATORIJA U EUROPSKOJ UNIJI

U Europskoj uniji jedanaest glavnih programa čini mrežu za provjeru uspješnosti laboratorija u analizi opće ili profesionalne izloženosti kemikalijama i prašinama u zraku. Usporedba ovih programa biti će prikazana u Zborniku programa koji se upravo sastavlja i omogućit će laboratorijima, korisnicima usluga i regulacijskim tijelima da izaberu program koji im najviše odgovara. Svi programi uspoređeni su s izmijenjenim Uputstvima ISO 43 (ISO Guide 43) objavljenim 1997. Statistika uspješnosti pokazuje da je većina programa u skladu s kriterijima Europskog standarda EN 482 koji utvrđuje granicu prihvatljivosti ukupne nepouzdanosti mjerenja. Statistika uspješnosti i strategije procjene uspješnosti razlikuju se od programa do programa.
Dok mnogi programi dostavljaju laboratorijima slične uzorke kao olovo na filtrima i benzen na ugljenu, mnogi bi programi htjeli uvesti i mnoge druge tipove uzoraka. To se, međutim, čini ekonomski neopravdanim na državnoj razini, tako da se unutar mreže razvija postupak kojim bi se ovi uzorci distribuirali svim zemljama članicama mreže. K tome, zemlje koje trenutno nemaju vlastiti program, a htjele bi ga uvesti moći će se obratiti mreži za pomoć pri uspostavljanju programa u tom području.

Ključne riječi:
analiza zraka, monitoring zraka, kemikalije, prašine, analiza okoliša, Europska mreža programa, osiguranje kakvoće, analiza radnog mjesta

Requests for reprints:

Peter R. Stacey, Ph.D.
The Health and Safety Laboratory
Broad Lane, Sheffield, S3 7HQ, United Kingdom
E-mail: *Peter_Stacey@hsl.gov.uk*